

Complex networks with scale-free nature and hierarchical modularity

Snehal M. Shekatkar and G. Ambika

Indian Institute of Science Education and Research, Pune 411008, India

Received: date / Revised version: date

Abstract. Generative mechanisms which lead to empirically observed structure of networked systems from diverse fields like biology, technology and social sciences form a very important part of study of complex networks. The structure of many networked systems like biological cell, human society and World Wide Web markedly deviate from that of completely random networks indicating the presence of underlying processes. Often the main process involved in their evolution is the addition of links between existing nodes having a common neighbor. In this context we introduce an important property of the nodes, which we call mediating capacity, that is generic to many networks. This capacity decreases rapidly with increase in degree, making hubs weak mediators of the process. We show that this property of nodes provides an explanation for the simultaneous occurrence of the observed scale-free structure and hierarchical modularity in many networked systems. This also explains the high clustering and small-path length seen in real networks as well as non-zero degree-correlations. Our study also provides insight into the local process which ultimately leads to emergence of preferential attachment and hence is also important in understanding robustness and control of real networks as well as processes happening on real networks.

PACS. 89.75.Da Systems obeying scaling laws – 89.75.Fb Structures and organization in complex systems – 89.75.Hc Networks and genealogical trees

1 Introduction

Complex networks have become a key tool for understanding the behavior of complex systems from diverse fields like biology [1, 2, 3], ecology [4], technology [5, 6] and social sciences [7, 8]. Such networks are found to have interesting properties not possessed by their completely random counterparts and this indicates the presence of robust organizing principles behind their growth [9]. The important features of such networks are existence of scale-free nature [10], high clustering and small average path length [11], hierarchical community structure or hierarchical modularity [12, 13, 14, 15] and non-zero degree correlations [16]. Over the years, various generative models of network growth have been proposed to explain one or more of these features [9, 17].

In scale-free network, the distribution of degrees follows a power law and this is usually attributed to preferential attachment [10]. The scale-free nature also implies small value for the average path length of the network [18]. Apart from being scale-free, most real networks are highly clustered as compared to completely random networks like Erdős-Rényi model [9]. This means that in these networks, two nodes which share a neighbor have quite high chance of themselves being connected to each other. This tendency can be quantified using either the global clustering coefficient C [19] or the Watts-Strogatz clustering coefficient

C_{WS} [11] of the network. Several variants of the basic preferential attachment model of Barabási and Albert (BA model) have been proposed to explain this high value of clustering along with scale-free structure [20]. In addition to scale-free structure, real networks also show modular organization of clustering. In such networks, there exist groups of nodes such that the nodes in the same group are very densely connected to each other whereas these groups themselves are connected to each other relatively sparsely. Such groups are known as communities in the network and the network is said to possess community structure or modular structure [21]. Existence of such modular structure implies high value of clustering coefficient C while the reverse is not true. Also, in many of these modular networks, modularity is hierarchical. This means that on a global scale, network is divided into communities of various sizes and each of those communities itself shows a modular structure inside it [12, 13, 14, 15].

It is understood that, in real networks, local processes are responsible for the emergence of overall global structure of the network [22]. The present understanding is that the preferential attachment should be regarded as an outcome of system-dependent local processes between the nodes, rather than a mechanism in itself [23, 24]. For example it has been suggested that gene duplication could be responsible for the scale-free structure of protein interaction networks [25]. A possible mechanism reported for the

local interactions which could produce scale-free networks with high clustering, is triadic closure where new links are added between the neighbors of a particular node [23, 26, 20, 27]. The models based on this mechanism can also give rise to simple or hierarchical community structure [23, 27]. In these models of triadic closure only new nodes are considered as sources of triadic closure. However, in a very general sense, in real networks, many new links are being created between existing nodes also. We feel the possibility of such “internal links” also should form an important part of the growth of such networks.

In the present paper we identify a fundamental property of hubs of complex networks that is almost system independent. We show that this property can potentially explain the simultaneous existence of scale-free structure and hierarchical modularity even with inclusion of internal links. Our model for growth of complex networks is also based on the concept of triadic closure but with an additional property for nodes called mediating capacity. This mediating capacity is a degree dependent property such that high degree nodes are weak mediators of triadic closure. As we argue in the following sections, hubs in any real network have capability to start and maintain interactions with diverse types of nodes. This makes the triadic closure around the hubs less likely and hence results in low value for their clustering coefficients endowing the network with hierarchical modularity.

The rest of the paper is organized as follows. Section 2 describes the existence of triadic closure in various real world networks along with the social networks. In Section 3 we introduce the concept of mediating capacity. Section 4 describes our model for growth of complex networks called mediated attachment model. In Section 5, we show the existence of hierarchical community structure in this model along with scale-free structure. Section 6 deals with time/size dependence of various important characteristics of the mediated network. Finally, in section 7 we draw conclusions related to mediated attachment models and indicate future directions of study.

2 Existence of triadic closure in complex networks

The triadic closure is proposed as one of the local processes that seem to be present in many real networks [20, 27]. This involves the formation of links between a pair of nodes of the networks which share one or more neighbors in the network (See Fig. 1(a)). We now describe the existence of mechanisms analogous to triadic closure in various real networks.

In the case of social networks like acquaintance, friendships, coauthors or Facebook networks, triadic closure is obviously present since usually people get to know each other through a common neighbor in the network. Languages are also viewed as a complex network of words which are connected to each other if they frequently occur in the same sentence [28, 29]. During the evolution of a particular language, when words A and B start appearing together in a sentence and if the same happens for

words B and C , with a fairly high probability words A and C would appear in the same sentence. In the case of web pages, most web pages keep evolving by the addition and modification of their content and hence new hyperlinks are usually added to the page. Suppose a web page A has hyperlink to another page B , which itself contains hyperlink to page C , then A can know about C through B and can connect to C . The co-citations also work the same way where instead of hyperlinks, bibliographies are used to know and connect to second neighbors [23].

All the network growth models involving triadic closure models assume that no direct preferential attachment is present while network is growing and preferential attachment is only an outcome of local process of triadic closure [23, 24, 27]. Such triadic closure models have been proved to be extremely successful in producing highly clustered scale-free networks. However the models of triadic closure proposed so far usually consider only new node as the source of triadic closure [20, 23, 30, 27]. As noted earlier, in real networks majority of links are in fact created by old nodes. So we have to include the formation of such internal links in the growth model. This requires the introduction of a new property of hubs which makes them weak mediators such that the generated network are scale-free and hierarchically modular.

3 Hubs as weak mediators

We now make an important observation about the nodes of all complex networks, especially scale-free networks. Usually hubs in scale-free networks are considered as most important nodes. We claim that the high degree nodes or hubs in complex networks have capability to interact with nodes of a wide range of properties. If we consider Google web page which is a hub in world wide web, the web pages which link to it are of diverse nature. Similarly, a hub in social network usually has interest in wide variety of topics and hence develops the connections to people from highly diverse fields. As our third example, we consider a network of words of a particular language. In this network, hub is usually a word that capable of appearing with words belonging to diverse range of contexts. For example, the word “THE” in English can appear with many different words. In the jargon of network theory, we can thus say that hubs are usually connected to dissimilar nodes. On the other hand, a low degree node (a Wikipedia page about a technical topic from mathematics) would have links only to nodes similar to it and hence usually its neighbors are also similar with each other.

In all these cases the probability that such diverse or dissimilar nodes will get connected is very low. This leads to the observation that hubs should function as weak mediators in such networks. For example, a hub in social network may introduce two of his contacts with each other but those two contacts would usually share different interests and hence the link may not get established between them. This implies that the mediating capacity of a node decreases with increase in degree. As we will see, this in-

sight naturally unifies the emergence of hierarchical community structure with scale-free nature in such networks.

4 Mediated attachment model

We present a generative mechanism in the growth of networks with the property of mediating capacity for the nodes as discussed in the previous section. To incorporate this, we conjecture that the average number of links mediated by a node with degree k decreases with k . Therefore we take the probability that a node i with degree k_i connects two of its neighbors at a given time to be $1/k_i^n$ where n is a parameter. For a node of degree k , maximum ${}^kC_2 \sim k^2$ links are possible among its k neighbors. Hence a node with degree k will form approximately $k^2(1/k^n) = (1/k^{n-2})$ links at any given time. Thus we must have $n > 2$ if hubs are to behave as weak mediators. We also describe the effect of $n < 2$ (i.e. the effect of hubs not being weak mediators) later in the section.

Since real networks operate locally, usually what happens in one part of the network doesn't directly affect the processes like addition of links and nodes in the other part of the network. This means that both the number of links and number of nodes in the network that join the network is an increasing function of the size of the network. To incorporate this type of increase in the number of nodes, at each discrete time t we add approximately $pN(t)$ number of nodes to the network where p is the parameter with value between 0 and 1 and $N(t)$ is the number of nodes in the network at time t . The case of addition of one node at a time would then correspond to p value that decreases with time.

The growth of the network in this model thus involves two processes:

1. Starting from two connected nodes at time $t = 0$, $\lceil pN \rceil$ number of nodes are added to the network at each discrete time t where $\lceil \cdot \rceil$ denotes the ceiling function. Each new node connects to one of the old nodes randomly with uniform probability.
2. At the same time, every existing node i of degree $k_i(t)$ (> 1) connects every pair of its neighbors with probability:

$$w_i(t) = \frac{A}{k_i(t)^n} \quad (1)$$

where A and n are adjustable parameters with $n > 2$. The quantity $w_i(t)$ reflects the mediating capacity of node i at time t .

A small network generated under these processes is shown in Fig. 1(b).

It can be easily seen that since each node is allowed to connect its neighbors with each other, the number of new links added to the network is an increasing function of the network size. If the rate of increase of network size is low, this would produce a highly dense network whose average degree $\langle k \rangle$ would increase with time. However, in the present model, number of nodes also increases quite fast

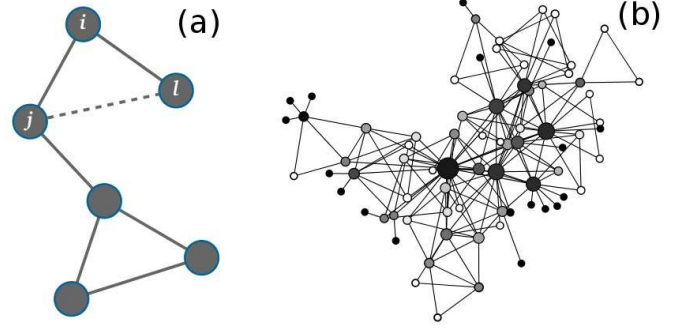


Fig. 1. (a) A graphical representation of triadic closure process: node i acts as a mediator for nodes j and l and tries to connect them. (b) Mediated network at time $t = 25$ with $p = 0.1$, $A = 8$ and $n = 3$. The sizes of nodes are proportional to their degree values and the color of each node is graded according to its clustering coefficient, the darkest being the lowest.

and as we will see below, these two factors can stabilize $\langle k \rangle$ of the network.

For the case of $n < 2$ we numerically compute the average degree $\langle k \rangle$ as a function of time and find that it increases monotonically. This implies high link density and non-stationarity of the degree distribution. In general $\langle k \rangle$ of the network must stabilize if the degree distribution is to be stationary asymptotically. In Fig. 2, $\langle k \rangle$ is plotted as a function of time for different values of n . The values for $n = 1.5$ are divided by 5 for better visualization. Clearly, for $n < 2$, $\langle k \rangle$ is a very rapidly increasing function of time and hence the degree distribution is not stationary in this case whereas for $n > 2$, it always stabilizes leading to a stationary degree distribution.

In Fig. 2 we show $\langle k \rangle$ as a function of time for different values of n for networks grown for 75 units of time. With $p = 0.1$, the results shown are for 100 realizations of the growth process. Fig. 3 shows the behavior of average degree as a function of parameter n at two different times. The rapid increase in average degree with time (and hence the existence of non-stationary degree distribution) is evident. This behavior is understandable because for small n , the process of creation of links overtakes the process of creation of nodes. For the mediating capacity chosen as in Eq.(1) from $n = 2$ onwards these two effects balance each other stabilizing $\langle k \rangle$.

Also when $n < 2$, the network is so dense that there are too many nodes with very high degree and hence the network does not exhibit a scale-free structure. To see this, we plot the degree distributions of the network for $n = 1.5$ and $n = 2.5$ in Fig. 5. For $n = 1.5$, there exists a peak in the tail and hence the network in this case is not scale-free.

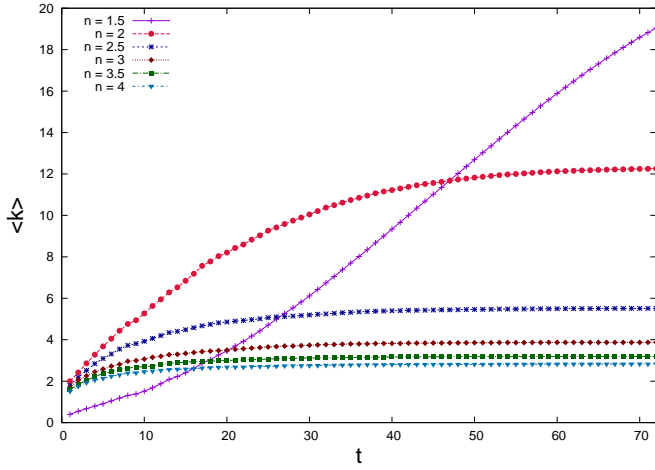


Fig. 2. (Color online) Average degree of the network as a function of time for various values of n with $A = 4$ and $p = 0.1$. For better visualization, for $n = 1.5$ the plotted values are equal to actual values divided by 5. For $n < 2$, $\langle k \rangle$ is a rapidly increasing function of time and hence leads to a non-stationary degree distribution. All cases with $n \geq 2$ lead to a stabilized averaged degree. In each case, network is grown up to size $N = 9525$ and results are averaged over 100 realizations.

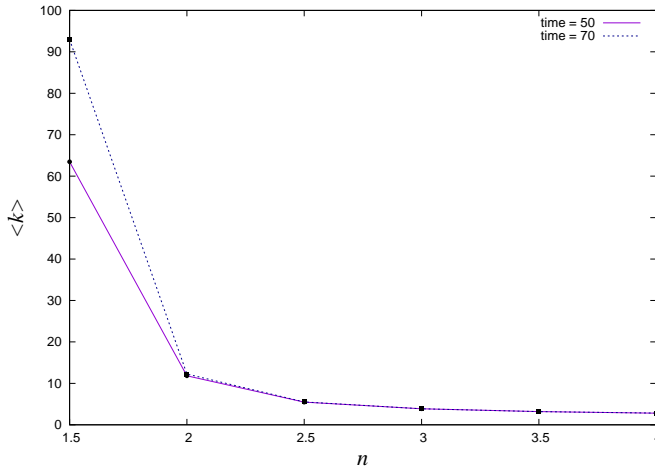


Fig. 3. (Color online) Average degree $\langle k \rangle$ as a function of parameter n at two different times $t = 50$ (a continuous line) and $t = 70$ (a solid line). It can be seen that when $n < 2$, $\langle k \rangle$ has tendency to increase rapidly while for $n > 2$ average degree saturates. The results are averaged over 100 random realizations.

5 Simultaneous occurrence of scale-free structure and hierarchical modularity

Many real networks are both scale-free and hierarchically modular. The simultaneous occurrence of scale-free structure and hierarchical modularity is a less explored topic in the context of generative network models. As we now show, the mediated attachment model based on the conjecture that the hubs are weak mediators naturally gives rise to networks with both these properties. To establish the scale-free structure of the resulting network, in Fig. 5

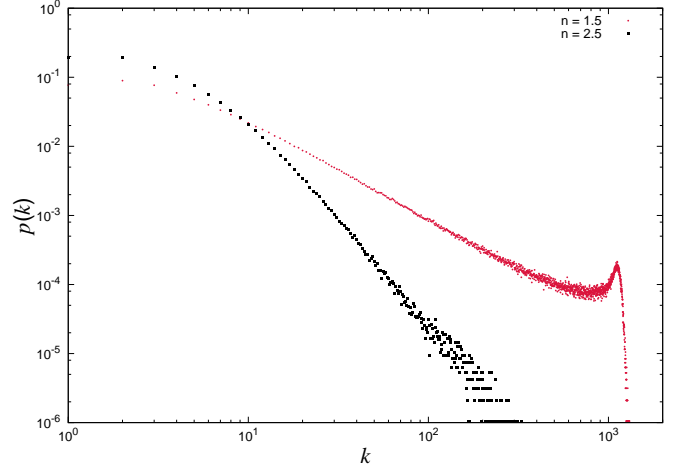


Fig. 4. (Color online) Comparison of degree distributions for mediated network with $n < 2$ and $n > 2$. Here for $n < 2$ ($n = 1.5$ in this case), the distribution completely deviates from the power law in the tail and is also non-stationary as shown in Fig. 2. Both cases correspond to $N = 9525$ and results are averaged over 100 realizations.

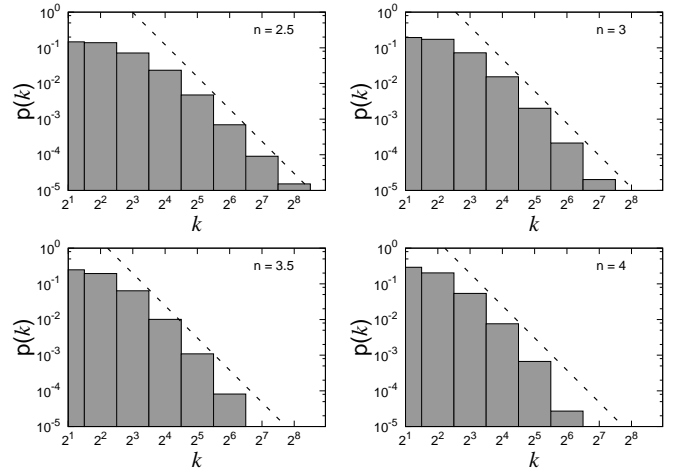


Fig. 5. Degree distributions of mediated network of size $N = 9525$ for different values of n with $A = 8$ and $p = 0.1$. A logarithmic binning is used to plot the histograms. The dotted line in each plot has slope -3 which is close to the actual scaling indices of real networks. All the results are averaged over 1000 random realizations.

we plot the degree distributions of the network for various values of n with $n > 2$ and all of them can be seen to follow a power-law behavior with a scaling index close to 3. The real networks are usually seen to have scaling indices in the range $(2, 4)$ [9,17] and hence the model can be tuned to produce the actually observed distributions satisfactorily.

A necessary (but not sufficient) indicator of existence of hierarchical modularity in the given network is the dependence of local clustering coefficient on the degree. A local clustering of the node i is given as:

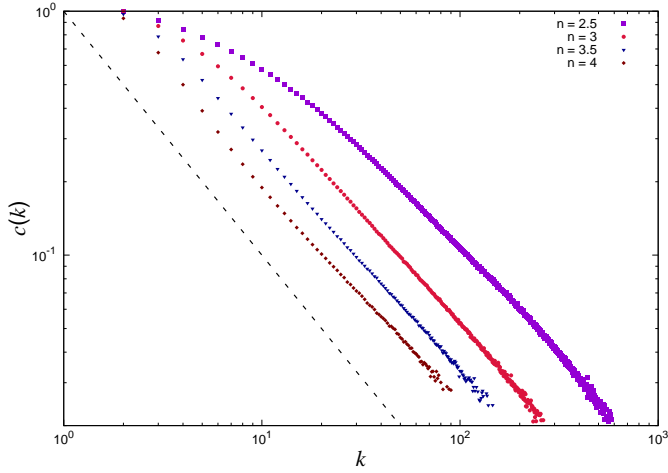


Fig. 6. (Color online) Local clustering coefficient c as a function of degree k for mediated network. The parameters are same as for Fig. 5. The black dotted line has slope -1 . Results are averaged over 1000 random realizations.

$$c_i = \frac{E_i}{k C_2} \quad (2)$$

where k is the degree of node and E_i is the number of links present among the neighbors of node i . Its value always lies between 0 and 1. For the networks with hierarchical modularity, c_i shows a systematic decrease with degree k with $c(k) \sim k^{-1}$ [13]. For the model presented here, we compute the local clustering coefficient for different values of n and plot it as a function of k in Fig. 6. The decrease of $c(k)$ with degree is clear in all cases. This happens because the mediating capacity of a node decreases with increase in its degree and which in turn reduces its clustering coefficient when degree is large.

To confirm that the mediated network does possess the hierarchical modularity, we use hierarchical Infomap (<http://www.mapequation.org/code.html>). We find that for all $n > 2$, the mediated network is hierarchically modular but the number of modular levels depends on the value of n and in fact is the increasing function of n . Thus, for the network size $N = 9525$ used here, for $n = 2.5$ network has 4 to 5 levels of hierarchy while for $n = 4$, the network has 8 to 9 levels of hierarchy.

6 Characteristics measures and their scaling

Now we look at the various important network characteristics like clustering coefficient, path length and degree correlations for the mediated network.

6.1 Clustering coefficient

The density of triangles in the network is quantified by the quantity called clustering coefficient. There exist two definitions of the clustering coefficient in the literature:

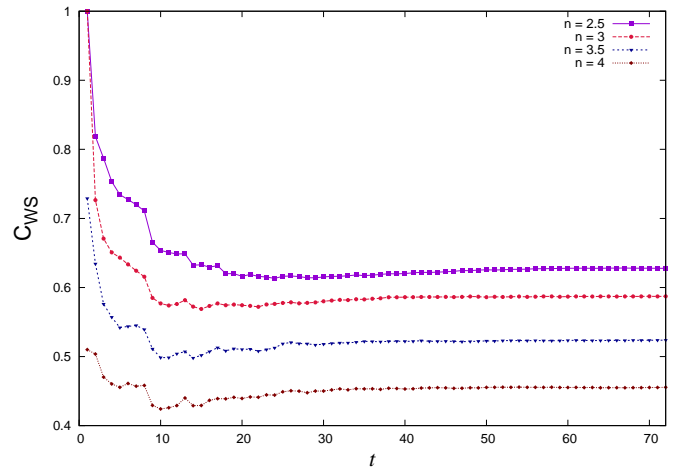


Fig. 7. (Color online) Watts-Strogatz clustering coefficient C_{WS} of the network as a function of time for various values of parameter n . All the parameter values are same as in Fig. 6

global clustering coefficient and Watts-Strogatz clustering coefficient. In the present case, we use Watts-Strogatz clustering coefficient C_{WS} (also called as average clustering coefficient) as the quantifier for the clustering in the network. It is defined as the average of the local clustering coefficients over all the nodes of the network:

$$C_{WS} = \frac{1}{N} \sum_{i=1}^N c_i \quad (3)$$

Fig. 7 shows the Watts-Strogatz clustering coefficient of the mediated network as a function of time for various values of n . It can be seen that all C_{WS} values settle asymptotically to values in the range (0.4, 0.7). Almost all the real networks like film actors, coauthorships, citations, Internet, WWW, metabolic networks, protein interactions, ecological food webs and neural networks show such high values of C_{WS} [31]. This agreement with the empirical observations further supports the model presented here.

6.2 Degree correlations

Apart from the properties discussed so far, many real networks show non-zero degree correlations. Networks in which high degree nodes tend to connect to low degree nodes (i.e. networks for which negative degree correlations exist) are known as dissortative networks while those in which similar degree nodes tend to connect to each other (i.e. networks for which positive degree correlations exist) are known as assortative networks.

Assortative or dissortative nature of networks can be quantified by calculating the assortativity coefficient for the network which is given by the following expression [17]:

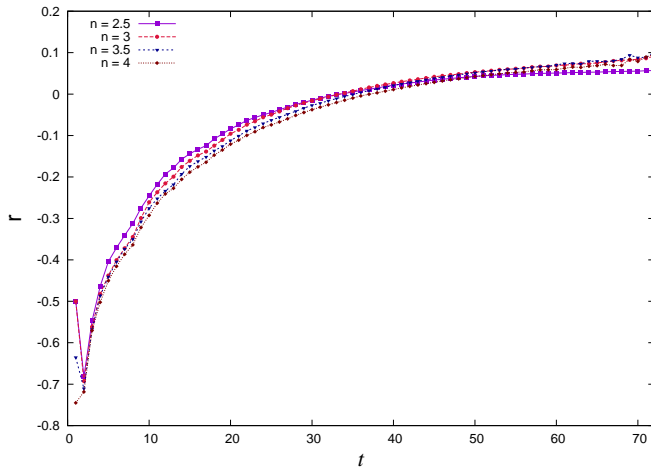


Fig. 8. (Color online) Assortativity coefficient r of the mediated network as a function of time for different values of parameter n . It can be seen that r stabilizes to positive values asymptotically and hence the mediated network is assortative. All the parameter values are as in Fig. 6

$$r = \frac{\sum_{ij} (B_{ij} - k_i k_j / 2m) k_i k_j}{\sum_{ij} (k_i \delta_{ij} - k_i k_j / 2m) k_i k_j} \quad (4)$$

Here, B_{ij} is $(i, j)^{th}$ element of an adjacency matrix of the network, m is the number of links in the network, k_i is the degree of node i and δ_{ij} is the Kronecker delta whose value is 1 when $i = j$ and is 0 otherwise.

Interestingly, social networks are always seen to possess positive degree correlations i.e they are assortative [16]. In Fig. 8 we show the assortativity coefficient of the mediated network as a function of time for different n . It is clear that our model with $n > 2$ leads to positive degree correlations.

6.3 Average path length

The real networks are found to have a small-world property which means that the average path length $\langle l \rangle$ of these networks varies very slowly with size and also they have high clustering coefficient [11]. We also calculate the $\langle l \rangle$ of mediated network as a function of its size for different n . It is clear that $\langle l \rangle$ increases logarithmically with size (see Fig. 9). Such logarithmic scaling with size has been reported earlier for many real networks [9]. This along with the observed high clustering, makes the mediated network a small-world network.

7 Conclusion

Understanding the physical mechanisms which lead to the emergence of various structural properties seen in real networks is an important part of the network science. Since the discovery of small-world nature and scale-free

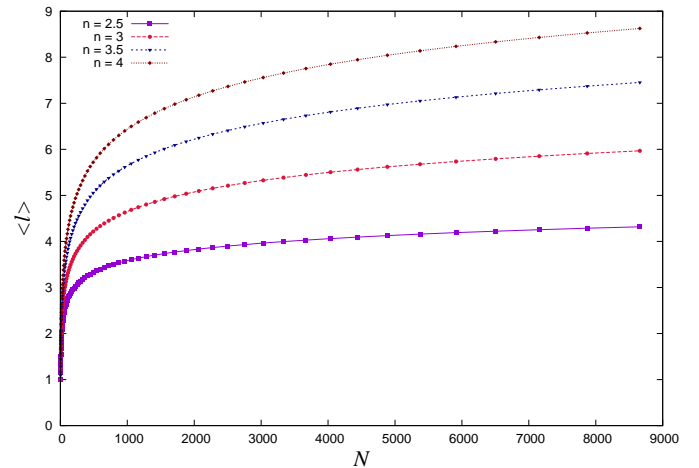


Fig. 9. (Color online) Average path length $\langle l \rangle$ of the network as a function of size $N(t)$ of the network for different n . The parameter values are as given in Fig. 6. Each curve in the plot follows a logarithmic increase.

nature of various real networks, better and better models of network formation have been proposed over the years. It has been observed that many real networks exhibit both scale-free nature and hierarchical modularity. However, the generic mechanisms which lead to their simultaneous occurrence in real networks are not yet very clear. In the present work we showed that the weakness of hubs in mediating the connections between their neighbors produces both of these structures successfully. The weakness of hubs can be understood from a more fundamental fact that hubs are usually connected to dissimilar nodes in the network. We feel that a more careful use of this idea in the context of specific networks will lead to many novel insights into the structure and function of various real networks.

The model presented here comes with tunable parameters p , A and n that can be adjusted to get desired values of network characteristics. For $n > 2$ we showed that the average degree of the network stabilizes and can be tuned to a desired value by tuning parameters A and n . Apart from scale-free nature and hierarchical modularity, we also showed that the mediated network is highly clustered and its path length scales logarithmically with size. This shows that a small-world nature is also an emerging property of this model. Finally, the network shows positive degree correlations as observed in case of social networks. We anticipate that detailed investigations to understand the intrinsic mechanisms that generate this property in different cases, can reveal interesting processes underlying the complexity of such systems.

One of the authors (S.M.S.) would like to thank University Grants Commission, New Delhi, India for financial assistance in the form of Senior Research Fellowship.

References

1. H. Jeong, B. Tombor, R. Albert, Z. Oltvai, A.L. Barabási, *Nature* **407**, 651 (2000)
2. H. Jeong, S. Mason, A.L. Barabási, Z. Oltvai, *Nature* **411**, 41 (2001)
3. J. White, E. Southgate, J. Thompson, S. Brenner, *Phil. Trans. R. Soc. London* **314** (1986)
4. J. Montya, R. Solé, *J. Theor. Biol.* **214**, 405 (2002)
5. R. Albert, H. Jeong, A.L. Barabási, *Nature* **401**, 130 (1999)
6. R. Pastor-Satorras, A. Vázquez, A. Vespignani, *Phys. Rev. Lett.* **87**, 258701 (2001)
7. M. Newman, *Phys. Rev. E* **64**, 016131 (2001)
8. M. Newman, *Phys. Rev. E* **68**, 036122 (2003)
9. R. Albert, A.L. Barabási, *Rev. Mod. Phys.* **74**, 47 (2002)
10. A.L. Barabási, R. Albert, *Science* **286**, 509 (1999)
11. D. Watts, S. Strogatz, *Nature* **393**, 440 (1998)
12. E. Ravasz, A. Somera, D. Mongru, Z. Oltvai, A.L. Barabási, *Science* **297**, 1551 (2002)
13. E. Ravasz, A.L. Barabási, *Phys. Rev. E* **67**, 026112 (2003)
14. A. Clauset, C. Moore, M. Newman, *Nature* **453**, 98 (2008)
15. C.A. Lewis, N. Jones, M. Porter, C. Deane, *BMC Systems Biology* **4**, 100 (2010)
16. M. Newman, *Phys. Rev. Lett.* **89**, 208701 (2002)
17. M. Newman, *Networks: An Introduction* (Oxford University Press, 2010)
18. F. Chen, Z. Chen, X. Wang, Z. Yuan, *Commun Nonlinear Sci Numer Simul* **13**, 1405 (2008)
19. M. Newman, S. Strogatz, D. Watts, *Phys. Rev. E* **64**, 026118 (2001)
20. P. Holme, B. Kim, *Phys. Rev. E* **65**, 026107 (2002)
21. M. Newman, *Nat. Phys.* **8**, 25 (2012)
22. R. Albert, A.L. Barabási, *Phys. Rev. Lett.* **85**, 5234 (2000)
23. A. Vázquez, *Phys. Rev. E* **67**, 056104 (2003)
24. T. Evans, J. Saramäki, *Phys. Rev. E* **72**, 026138 (2005)
25. A. Vázquez, A. Flammini, A. Maritan, A. Vespignani, *Complexus* **1**, 38 (2002)
26. J. Davidsen, H. Ebel, S. Bornholdt, *Phys. Rev. Lett.* **88**, 128701 (2002)
27. G. Bianconi, R. Darst, J. Iacovacci, S. Fortunato, *Phys. Rev. E* **90**, 042806 (2014)
28. R.F. i Cancho, R. Solé, *Proc. R. Soc. Lond. B* **268**, 2261 (2001)
29. R.F. i Cancho, R. Solé, R. Köhler, *Phys. Rev. E* **69**, 051915 (2004)
30. J. Saramäki, K. Kaski, *Physica A* **341**, 80 (2004)
31. M. Newman, *SIAM REVIEW* **45**, 167 (2003)